**3rd week VIVA QUESTIONS:**

# **What do you mean by data type casting in the context of data analysis and data cleaning?**

In the context of data analysis, casting data types means converting data from one type to another. This is often done to ensure consistency and accuracy in data analysis, as well as to enable specific operations or functions that are available for certain data types

1. When is it appropriate to drop missing values in data rather than imputing them in the

context of data cleaning with Pandas?

It is appropriate to drop missing values in data when the amount of missing data is small compared to the overall size of the dataset, and the missing data is randomly distributed or when they would skew the analysis

# What is Pandas in Python?

# Pandas is an open-source Python package that is most commonly used for data science, data analysis, and machine learning tasks. It is built on top of another library named Numpy

#### **What are the steps involved in data transformation?**

The steps involved in data transformation can vary depending on the specific transformation being performed, but typically involve extracting the data from its original source, converting it to the desired format, and then loading it into the new destination.

#### 

#### **Can you explain why data cleansing is important**

# 

Having clean data will ultimately increase overall productivity and allow for the highest quality information in your decision-making

**4th week VIVA QUESTIONS:**

#### **What are some of the features provided by Matplotlib?**

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update.

1. **What is the difference between a bar plot and a histogram**

A bar plot displays discrete data as bars, while a histogram displays continuous data as bars that represent the frequency of data points in a given range

1. **What is the difference between a scatter plot and a line plot**

A scatter plot displays individual data points as markers, while a line plot connects data points

with a line.

### 

### What is a Boxplot?

## 

## A Box and Whisker Plot, also known simply as a Boxplot, is a graphical representation of data distribution through its quartiles. The graph is presented in the form of a rectangle with lines emanating from both the top and the bottom. These lines, called "whiskers," depict the variability outside the upper and lower quartiles

#### 

#### **What’s the best way to draw multiple lines in a single figure?**

The best way to draw multiple lines in a single figure is to use the plot() function. This function can take multiple arguments, each of which will be plotted as a separate line. For example, to plot two lines, you would use the following code:

plot(x1, y1, x2, y2)

**5th week VIVA QUESTIONS**

## What is web scraping?

Web scraping is the technique to extract and read the data from the internet. The collected

data can be saved and reused for data analytics

1. **What is the purpose of the request module in Python?**

The request module is used to read the data from the internet web pages.You have to

pass the URL from where you want to read the data along with the HTTP request method,

header information like encoding method, response data format, and session cookies

## What is the difference between HTML parsing and XML parsing?

The main difference between HTML parsing and XML parsing is that HTML parsing is

designed to be forgiving of errors, while XML parsing is not.This means that if you are trying

to parse an HTML document, you will be able to still extract information even if the document

is not well-formed. However, if you try to parse an XML document that is not well-formed, you

will likely not be able to extract any information at all

1. **What is a robots.txt file?**

Robots.txt is a text file that tells crawlers, bots, or spiders if a website could be or how it

should be scrapped as specified by the website owner. It is critical to understand the robots.txt

file to prevent being blocked while web scraping.

## Is web scraping data mining?

Web scraping and data mining are two different concepts. Web scraping is to collect

raw data, but data mining is the process of discovering patterns in large data sets.

## 